Multidimensional Data Aggregation Scheme for Smart Grid with Differential Privacy

Xiuxia Tian[1,2], Qian Song[1] and Fuliang Tian[1]
(Corresponding author: Xiuxia Tian)

School of Computer Science and Technology, Shanghai University of Electric Power[1]
No. 2588 Changyang Road, Shanghai 200090, China
(Email: xxtian@fudan.edu.cn)
College of Data Science and Engineering, East China Normal University[2]
No. 3663 Zhongshan Road, Shanghai 200062, China
(Received Sept. 2, 2017; revised and accepted Nov. 28, 2017)

Abstract

The use of smart meters allows the power supplier to collect detailed consumer data from consumers, which may threaten consumers' personal information. In order to protect the privacy of consumers and prevent data leakage from specific consumers, we propose a multidimensional data aggregation scheme with differential privacy. The proposed scheme uses the Horner rule to deal with multidimensional data. The proposed scheme uses certificate-based aggregate short signature to achieve data authentication and data integrity, which reduces the number of bilinear pairing to a constant. Specially, our proposed scheme overcomes the differential attack problem by adding Laplace noise to aggregated data. We analyze the level of differential privacy utility. Compared to existing schemes, the proposed scheme is more efficient in terms of computational cost and communication overhead.

Keywords: Data Aggregation; Differential Privacy; Privacy Protection; Smart Grid

1 Introduction

In the past few years, we have seen increasing interest in smart grid technology around the world [2]. Smart grid, or future power grid, which is a combination of traditional grid systems and advanced information and communication technologies, improves the efficiency, reliability, economy and power generation continuity of the modern grid and provides more stable and reliable power for power users [19, 25]. Advanced communications and information technology applications solve many inherent problems in traditional grids, such as lack of load balancing, intelligent consumption and dynamic pricing [9]. Smart meters in the smart grid system collects consumer power consumption data and other information and send it to the remote control center [11, 20]. The widespread deployment of smart meters has also brought problems about privacy leaks [13]. Smart meters store consumer sensitive power consumption information because they can be used to analyze the user's lifestyle [10]. During the data transmission process, the authenticity, integrity and availability of the data may be destroyed, the user's personal sensitive information may be attacked by the attacker.

Varieties of security protection technologies have been developed in smart metering [21]. The privacy protection of existing smart grid communication process is mostly based on data aggregation technology. The user encrypts the usage data and sends it to the gateway. The gateway authenticates all the received data and sends them to the operation center after they are aggregated. Data aggregation technology refers to the use of encryption methods to enable power companies to calculate the total power of all users to analyze the data without knowing the power consumption of each user [25]. Existing data aggregation schemes use techniques such as homomorphic encryption [15, 16, 17, 19], blind factors [8] and differential privacy [3, 4, 18]. However, many of these schemes can not achieve both integrity and confidentiality. Also they can not resist some specific attacks. At the same time, many studies are concerned about the one-dimensional data, but with the development of smart meter technology we need to focus on dealing with multidimensional data.

In order to solve the existing problems of data aggregation scheme, we propose a multidimensional data aggregation scheme with differential privacy.

The contributions of this paper are in the following:

1) The proposed scheme uses certificate-based aggregate short signatures [12]. The aggregated signatures used in the scheme are short and efficient, which reduces the number of bilinear pairing to a constant. The security analysis demonstrates we can achieve data

authentication and data integrity.

2) To resist the differential attack, the proposed scheme provides ε-differential privacy by adding the Laplace noise selected from the Laplace distribution to the aggregated data of the community gateway.

3) Compared with others schemes, the proposed scheme has lightweight computational cost and communication overhead.

The remainder of this work works as follows: Section 2 describes the work of the existing data aggregation scenario. Section 3 describes the system model and the security model. Section 4 outlines the relevant preliminaries. Section 5 presents our multidimensional data aggregation scheme. Section 6 and Section 7 give the security and performance analysis. Finally we draw our conclusions in Section 8.

2 Related Works

In order to solve the privacy problem in the smart grid data aggregation, a variety of data aggregation privacy protection scheme are proposed. Li *et al.* [15] proposed a method of incremental aggregation in the network using homomorphic encryption, which did not solve the problem of authentication and data integrity. Li *et al.* [16] proposed a scheme about privacy protection demand response, which is based on homomorphic encryption and identity-based signature to achieve the security of the one-dimensional data aggregation. It also used adaptive key evolution technology for demand response. Bao *et al.* [3] proposed a lightweight data aggregation scheme that added symmetric geometric noise to resist differential attacks and achieved fault tolerance. The scheme used non-interactive session keys for source authentication and integrity protection.

Chen *et al.* [5] proposed a multifunctional data aggregation scheme that implemented statistical functions for usage data, such as averaging, variance and so on. Fan *et al.* [8] proposed the first one-dimensional data aggregation solution for internal attackers, which utilized blind factors to process confidential data and used small indexes to improve batch validation to achieve security utility. He *et al.* [10] improved the key leak problem in Fan's scheme, by reducing the number of bilinear pairing operations. In order to reduce the computational cost, He *et al.* [11] continued to improve the scheme, by using elliptic curve cryptography and implementing a lightweight data aggregation based on the *Schnorr* signature scheme. Abdalla *et al.* [1] used the $NTRU$ cryptosystem to achieve privacy protection, and the new ring signature NSS was signed to ensure integrity. But the scheme focused on the prediction of the demand for electricity from a group of customers in the same region instead of focusing on the process of data aggregation.It also restricted the connection with the provider only when the total clusters demand needed to be adjusted.

All the aforementioned schemes consider one-dimensional data aggregation. In order to deal with the problem of multidimensional data aggregation, Lu *et al.* [17] proposed the use of super-increasing sequence to process multidimensional electricity data. The *Paillier* homomorphic encryption system was used to encrypt the aggregated data during data transmission, and they used BLS short signature and batch verification. But the scheme had only one residential area and a gateway, which limited the size of the user and could not resist the differential attack. Fu *et al.* [9] used the elliptic curve *ElGamal* cryptography to encrypt multidimensional data and adopted the method of aggregated signature to carry out secure multidimensional data aggregation. However, the data aggregation result obtained by this scheme was the sum of all data. They could not separate the fine-grained data of each dimension. Zhou *et al.* [26] proposed a multilevel network aggregation scheme with fault tolerance and invalid signatures search. Shen *et al.* [19] used the Horner rule to handle multidimensional data by using two-level gateways to protect privacy. The aggregation scheme could handle dynamic users, but it could not resist differential attacks.

3 System Model And Security Requirements

In this section we will introduce the system model and security requirements.

3.1 System Model

The system is a four-layer smart grid communication network structure, which consists of the operation center, regional gateway, community gateway and home area network (*ie* users). The system model is shown in Figure 1. An operation center is responsible for a region, corresponding to the regional gateway; a region has m communities, corresponding to the community gatewayBGW_1, BGW_2,BGW_3...BGW_m. The first i community has n_i users, and each user collects their electricity data with a smart meter.

- OC: OC is a trusted entity that is responsible for the registration verification of the gateway and the user. It issues certificates for all gateway and users, generates keys for the entire system and issues public parameters. It also collects, processes and analyzes real-time data, such as the sum of power usage data for a given dimension or the power consumption during peak hours, which implements segmented power pricing decisions and appropriate resource allocation to provide reliable service for smart grid systems.

- DGW: DGW has the function of aggregation and relay. It's responsible for verifying messages from various BGWs and aggregating them. Then DGW forwards the messages to OC.

- *BGW*: *BGW* has the same function as *DGW*, responsible for verifying the confidential information received by the users. Then *BGW* aggregates the information and forwards them to the higher level gateway. In this process, *BGW* may be easily attacked by external attackers (such as differential attacks) because of the low level of security.
- *U*: The smart meter owned by the user is responsible for periodically (for example, every 15 minutes) to encrypt the coarse-grained l-dimensional data and report it to *BGW*.

The communication between the user and *BGW* generally uses *WiFi* technology, while *BGW* and *DGW* , *DGW* and *OC* generally use the wired link to communicate.

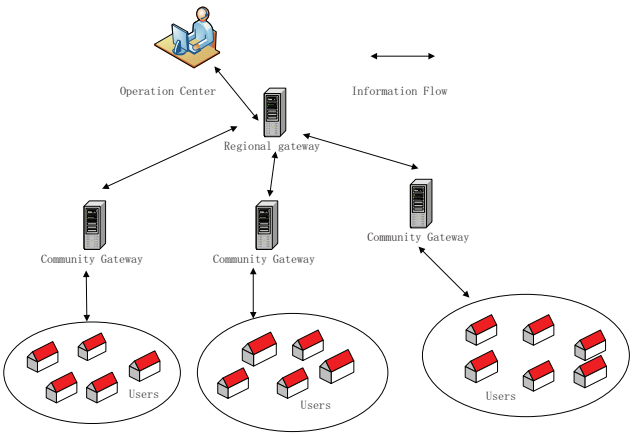

Figure 1: The proposed scheme

3.2 Security Requirements

In this system, we believe that *OC*, *DGW*,*BGW* are completely trusted, but users are semi-honest, which means users will not deliberately leak or change the information, but they are curious about others, trying to infer the usage information of others.

There may be an adversary A in the system who will steal the usage data when the user sends their data to the gateway; or an adversary attack (such as a differential attack) by analyzing the similar data sets obtained from the two aggregations may happen, trying to infer the individual user's sensitive information from the aggregation difference.

Adversary A may also invade the database of *DGW* and *BGW*, or invade communication links, which will destroy the authenticity and integrity of the data.

Confidentiality: Even if some users may collude with each other,they can not get the usage information of other users. Adversary A who steals the electricity data can not get the relevant information of a single user.

Authentication and data integrity: The user's electricity data in the transmission process requires authentication to avoid being tampering or forgery by malicious attackers.

Differential privacy: Even if the opponent A launches a differential attack, he/she can not get the individual user's sensitive electricity data.

4 Preliminaries

In this section we will outline the knowledge of Bilinear pairing, *Paillier* encryption algorithm, horner rule, and differential privacy as the basis of our scheme.

4.1 Bilinear Pairing

Let k be the security parameter and p be the prime with k bits. Let G_1,G_T be a cyclic addition group of order p generated by P (generator); a and b are elements in Z_p (Z_p is the prime p-order cyclic group).

Assume that the discrete logarithm problem in G_1 and G_T is a difficult problem. Bilinear pairing is a mapping that satisfies the following properties [22]:

- Bilinearity: For any $P, Q \in G_1$ and $a, b \in Z_p$, we have $e(aP, bQ) = e(P, Q)^{ab}$.
- Non-degenerative: There are $P, Q \in G_1$, making $e(P, Q) \notin 1$.
- Computability: For all $P, Q \in G_1$, there exists valid algorithms to calculate $e(P, Q)$.

4.2 Paillier Encryption Algorithm

In the *Paillier* cryptography system, the public key is $pk(N, g)$, the corresponding private key is $sk(\lambda, \mu)$. Let $E(\cdot)$, m and r represent encrypted functions, messages and random numbers respectively. The ciphertext is $c = E(m) = g^m \cdot r^N mod N^2$. The plaintext is $m = D(c) = L(c^{\lambda mod N^2}) \cdot \mu mod N$.

4.3 Horner Rule [19]

The Horner rule uses the least multiplication strategy to find the value of the polynomial $A = a_n x^n + a_{n-1} x^{n-1} + ...+a_1 x+a_0$ at x. Using this rule, n coefficients a_1,a_2,...,a_n are obtained by n multiplications and n additions.

4.4 Differential Privacy [6]

If the data aggregation result D includes the user Bob's power usage data, the algorithm M (D) is executed to obtain some calculation results. Assuming that the data aggregation result D is changed to D' after deleting of Bob's data , the execution of the algorithm M (D') or M (D) produces almost the same result. It is assumed that Bob's usage data is safe in the data set D under the algorithm M. It means whether Bob's data exists or not will not affect the output.

4.4.1 ε-Differential Privacy

If the data sets $D_1$and D_2 are different for at most one element, the randomization function K gives ε- differential privacy, that is, for any $s \in Range(k)$, we have $P[k(D_1) \in s] \leq e^{\varepsilon} P[k(D_2) \in s]$.

4.4.2 Laplace Mechanism [7]

Laplace mechanism is to use the Laplace distribution to produce noise. The probability density function of the Laplace distribution is $p(x) = \frac{1}{2b} e^{-\frac{|x|}{b}}$. From the Laplace distribution, noise r is randomly selected to be added to the original aggregation, the perturbed result can achieve ε- differential privacy.

5 Multidimensional Data Aggregation Scheme

In this paper, the multi-dimensional data aggregation scheme with differential privacy is divided into five stages: initialization phase, registration phase, user data encryption phase, secure aggregation phase and data recovery phase.

The processing of each stage is described in detail below and the meaning of each character symbol is shown in Table 1.

5.1 Initialization

At this stage, OC generates and publishes the parameters. Giving the security parameters k and k_1, OC guides the entire system.

Step 1. OC generates (q, P, G_1, G_2, e) by running $Gen(k)$. Then OC calculates the $Paillier$ encryption system's public key $(N = pq, g)$ and the corresponding private key (λ, μ).

Step 2. Define four secure hash functions $H_0 : \{0,1\}^* \to Z_q^*$, $H_1 : \{0,1\}^* \to G_1$, $H_2 : \{0,1\}^* \to Z_q^*$, $H_3 : \{0,1\}^* \to G_1$.

Step 3. Select two random factors R and R_1. After choosing the random number $s \in Z_q^*$ as the master key, OC compute the relevant public key $P_{pub} = sP$ and $Q_{pub} = H_1(P, P_{pub})$.

Step 4. Finally, OC publishes the system public information $\{q, P, G_1, G_T, e, N, g, R, R_1, H_0, H_1, H_2, H_3, P_{pub}, Q_{pub}\}$.

5.2 Registration Phase

In this section, we have to complete the registration of regional gateways , community gateways and users to ensure their legitimacy. The registration phase includes regional gateway registration, community gateway registration and user registration.

5.2.1 Regional Gateway Registration

DGW registers in OC and OC issues a certificate for it. After registration, DGW becomes a legal regional gateway. The process is as follows.

Step 1. DGW generates two random numbers $r, x \in Z_q^*$, calculates $X = xP$, $\alpha = rP$ and $\beta = r - xH_0(ID \parallel \alpha) mod q$ where is the x private key, X is the public key. DGW then sends the message$\{X, \alpha, \beta, ID\}$to the OC.

Step 2. OC checks the equation $\alpha = P\beta + XH_0(ID \parallel \alpha)$. If it does not exist, OC will deny the registration; otherwise, OC calculate $Q_{ID} = H_3(ID, X)$ and issues a certificate $cert_{ID} = sQ_{ID}$.

Verification of regional gateway registration:

$$\begin{aligned} & P\beta + XH_0(ID \parallel \alpha) \\ =& P(r - xH_0(ID \parallel \alpha) mod q) + XH_0(ID \parallel \alpha) \\ =& rP - xPH_0(ID \parallel \alpha) + XH_0(ID \parallel \alpha) \\ =& \alpha - XH_0(ID \parallel \alpha) + XH_0(ID \parallel \alpha) \\ =& \alpha \end{aligned} \tag{1}$$

5.2.2 Community Gateway Registration

BGW registers in OC and OC issues a certificate for it. After registration, BGW becomes a legal community gateway. The process is as follows.

Step 1. BGW_i generates two random numbers $r_i, x_i \in Z_q^*$, calculates $X_i = x_i P$, $\alpha_i = r_i P$ and $\beta_i = r_i - x_i H_0(ID_i \parallel \alpha_i) mod q$ where is the x_i private key, X_i is the public key. Then BGW sends the message $\{X_i, \alpha_i, \beta_i, ID_i\}$ to the OC.

Step 2. OC checks the equation $\alpha_i = P\beta_i + X_i H_0(ID_i \parallel \alpha_i)$. If it is not met, OC will deny the registration; otherwise, OC calculates $Q_{ID_i} = H_3(ID_i, X_i)$ and issues a certificate $cert_{ID_i} = sQ_{ID_i}$.

Verification of community gateway registration:

$$\begin{aligned} & P\beta_i + X_i H_0(ID_i \parallel \alpha_i) \\ =& P(r_i - x_i H_0(ID_i \parallel \alpha_i) mod q) + X_i H_0(ID_i \parallel \alpha_i) \\ =& r_i P - x_i P H_0(ID \parallel \alpha) + X_i H_0(ID_i \parallel \alpha_i) \\ =& \alpha_i - X_i H_0(ID_i \parallel \alpha_i) + X_i H_0(ID_i \parallel \alpha_i) \\ =& \alpha_i \end{aligned} \tag{2}$$

5.2.3 User Registration

U registers in OC and OC issues a certificate for it. After registration, U becomes a legal user. The specific process is as follows.

Step 1. U_{ij} generates two random numbers $r_{ij}, x_{ij} \in Z_q^*$, calculates $X_{ij} = x_{ij} P$, $\alpha_{ij} = r_{ij} P$ and $\beta_{ij} = r_{ij} - x_{ij} H_0(ID_{ij} \parallel \alpha_{ij}) mod q$ where the x_{ij} is private key, X_{ij} is the public key. Then U_{ij} sends the message $\{X_{ij}, \alpha_{ij}, \beta_{ij}, ID_{ij}\}$ to the OC.

Table 1: Notations

Notations	Description
OC	Operation Center
DGW	Regional Gateway
BGW	Community Gateway
U	User
$k, q, P,$	Bilinear pair parameter
k_i, p, q, g	Paillier password system parameters
H_0, H_1, H_2, H_3	Hash functions
R, R_i	The factor that handles multidimensional data
n	The maximum number of users in a community
n_i	The number of users in the i-th community
r	Random number
ε	Differential privacy budget

Step 2. OC checks the equation $\alpha_{ij} = P\beta_{ij} + X_{ij}H_0(ID_{ij} \parallel \alpha_{ij})$. If it is not met, OC will deny the registration; otherwise, OC calculates $Q_{ID_{ij}} = H_3(ID_{ij}, X_{ij})$ and issues a certificate $cert_{ID_{ij}} = sQ_{ID_{ij}}$.

Verification of user registration:

$$\begin{aligned} & P\beta_{ij} + X_{ij}H_0(ID_{ij} \parallel \alpha_{ij}) \\ =& P(r_{ij} - x_{ij}H_0(ID_{ij} \parallel \alpha_{ij}) mod q) + X_{ij}H_0(ID_{ij} \parallel \alpha_{ij}) \\ =& r_{ij}P - x_{ij}PH_0(ID_{ij} \parallel \alpha_{ij}) + X_{ij}H_0(ID_{ij} \parallel \alpha_{ij}) \\ =& \alpha_{ij} - X_{ij}H_0(ID_{ij} \parallel \alpha_{ij}) + X_{ij}H_0(ID_{ij} \parallel \alpha_{ij}) \\ =& \alpha_{ij} \end{aligned} \tag{3}$$

5.3 User Data Encryption

The user data encryption generation phase is responsible for handling users' multidimensional data. Users regularly collect their l-dimensional usage data$(d_{ij1}, d_{ij2}, ..., d_{ijl})$. Like [19], we handle multi-dimensional data by synthesizing the l-dimensional data into a polynomial and implements the following steps:

Step 1. Structure a polynomial with l-dimensional usage data $M_{ij} = R_i^i(d_{ij1}R^1 + d_{ij2}R^2 + ... + d_{ijl}R^l)$;

Step 2. The user selects r_{ij}^* and calculates the ciphertext $C_{ij} = g^{M_{ij}}r_{ij}^{*N} mod N^2$;

Step 3. Calculate $h_{ij1} = H_0(ID_{ij} \parallel C_{ij} \parallel X_{ij} \parallel T \parallel P_{pub})$,$h_{ij2} = H_2(ID_{ij} \parallel C_{ij} \parallel X_{ij} \parallel T \parallel Q_{pub})$,$V_{ij} = h_{ij1} \cdot cert_{ID_{ij}} + x_{ij} \cdot h_{ij2} \cdot Q_{pub}$. The signature is $\sigma_{ij} = (C_{ij}, V_{ij})$;

Step 4. Send $(ID_{ij}, \sigma_{ij}, C_{ij}, T)$ to BGW_i.

5.4 Secure Aggregation Phase

In this section we mainly complete the secure aggregation of data in the community gateway and the regional gateway. Before the gateway data aggregation, we rely on the aggregation of certificate-based short signatures to complete the security certification, which ensures the integrity of the data during transmission process. The length of the signature and the number of bilinear pairing involved in the algorithm are independent of the number of users. After the community gateway data is aggregated, we add noise to the community gateway to achieve differential privacy.

5.4.1 Community Gateway Aggregation

The community gateway aggregates the signatures of the received data and verifies it. Then the community gateway aggregates the encrypted data for all users in their own community. During this process, the community gateway will add Laplace noise to the encrypted data to resist differential attack. To add Laplace noise to the aggregated data, the sensitivity of the data set need to be calculated.

Let D be a subset of the users and for two data sets D_1 and D_2 with only one element different, we have $\| A(D_1) - A(D_2) \|_1 \leq W$. Therefore, the sensitivity of A is $\Delta f = W$.

Step 1. Verify the received data $(ID_{ij}, \sigma_{ij}, C_{ij}, T)$. Then BGW calculates received n_i users aggregate signature $V_0 = \sum_{j=1}^{n_i} V_{ij}$ by using the aggregate signature generator. The corresponding signature set is $\{(C_{i1}, V_{i1}), (C_{i2}, V_{i2}), ..., (C_{in_i}, V_{in_i})\}$.

Step 2. Check that if the verification $e(V_0, P) = e(\sum_{j=1}^{n_i} h_{ij1}Q_{ID_{ij}}, P_{pub})e(\sum_{j=1}^{n_i} h_{ij2}X_{ij}, Q_{pub})$ is met.

Step 3. If the verification is successful, the BGW aggregates $n = n_i + n - n_i$ encrypted l-dimensional data items $C_{i1}, C_{i2}, ..., C_{in}$, where the n_i usage data reports are received from smart meters and the remaining $n - n_i$ data reports are constructed from

the zero-dimensional vector. Then aggregate n ciphertexts $C_{GW_i} = \prod_{j=1}^{n} C_{ij} mod N^2$.

Step 4. According the sensitivity $\Delta f = W$, BGW generates a noise $\tilde{m}_i$ from the Laplace distribution. The final result is calculated as $\tilde{C}_{GW_i} = C_{GW_i} \cdot g^{\tilde{m}_i}$. The resulting ciphertext is the disturbed data and the noise therein ensures the privacy of each user.

Step 5. BGW calculates $h_{i1} = H_0(ID_i \parallel \tilde{C}_{GW_i} \parallel X_i \parallel T \parallel P_{pub})$, $h_{i2} = H_2(ID_i \parallel \tilde{C}_{GW_i} \parallel X_i \parallel T \parallel Q_{pub})$, $V_i = h_{i1} \cdot cert_{ID_i} + x_i \cdot h_{i2} \cdot Q_{pub}$. The signature is $\sigma_i = (\tilde{C}_{GW_i}, V_i)$;

Step 6. BGW sends $(ID_i, \sigma_i, \tilde{C}_{GW_i}, T)$ to DGW_i.

Verification of correctness:

$$\begin{aligned} & e(V_0, P) \\ =& e(\sum_{j=1}^{n_i} V_{ij}, P) \\ =& e(\sum_{j=1}^{n_i} (h_{ij1} cert_{ID_{ij}} + x_{ij} h_{ij2} Q_{pub}), P) \\ =& e(\sum_{j=1}^{n_i} h_{ij1} cert_{ID_{ij}}, P) e(\sum_{j=1}^{n_i} x_{ij} h_{ij2} Q_{pub}, P) \\ =& e(\sum_{j=1}^{n_i} h_{ij1} Q_{ID_{ij}}, P_{pub}) e(\sum_{j=1}^{n_i} h_{ij2} X_{ij}, Q_{pub}). \end{aligned} \tag{4}$$

5.4.2 Regional Gateway Aggregation

The regional gateway aggregates the signatures of the received data and verifies it. Then the regional gateway aggregates the encrypted data for all users in its region.

Step 1. For the received data $(ID_i, \sigma_i, \tilde{C}_{GW_i}, T)$, DGW uses the aggregate signature generator to calculate $V_0^* = \sum_{i=1}^{m} V_i$. Then DGW uses the aggregated short signature verification method to verify if $e(V_0^*, P) = e(\sum_{i=1}^{m} h_{i1} Q_{ID_i}, P_{pub}) e(\sum_{i=1}^{m} h_i X_i, Q_{pub})$.

Step 2. If the above verification is successful, the aggregation operation $\tilde{C} = \prod_{i=1}^{m} \tilde{C}_{GW_i} mod N^2$ is performed by DGW.

Step 3. DGW calculates $h_1 = H_0(ID \parallel \tilde{C} \parallel X \parallel T \parallel P_{pub})$, $h_2 = H_2(ID \parallel \tilde{C} \parallel X \parallel T \parallel Q_{pub})$, $V = h_1 \cdot cert_{ID} + x \cdot h_2 \cdot Q_{pub}$. The signature is $\sigma = (\tilde{C}, V)$;

Step 4. DGW sends $(ID, \sigma, \tilde{C}, T)$ to OC.

Verification of correctness:

$$\begin{aligned} & e(V_0^*, P) \\ =& e(\sum_{i=1}^{m} V_i, P) \\ =& e(\sum_{i=1}^{m} (h_{i1} cert_{ID_i} + x_i h_{i2} Q_{pub}), P) \\ =& e(\sum_{i=1}^{m} h_{i1} cert_{ID_i}, P) e(\sum_{i=1}^{m} x_i h_{i2} Q_{pub}, P) \\ =& e(\sum_{i=1}^{m} h_{i1} Q_{ID_i}, P_{pub}) e(\sum_{i=1}^{m} h_{i2} X_i, Q_{pub}). \end{aligned} \tag{5}$$

5.5 Data Recovery

For the received $(ID, \sigma, \tilde{C}, T)$, OC verifies if the signature $e(V, P) = e(h_1 Q_{ID}, p_{pub}) e(h_2 X, Q)$. After successful verification, the $Paillier$ decryption algorithm is used to obtain the sum of all multidimensional data. Then We use Horner rules to analyze the sum of the data for each dimension.
Verification of correctness:

$$\begin{aligned} & e(V, P) \\ =& e(h_1 cert_{ID} + x h_2 Q_{pub}), P) \\ =& e(h_1 cert_{ID}, P) e(x h_2 Q_{pub}, P) \\ =& e(h_1 Q_{ID}, P_{pub}) e(h_2 X, Q_{pub}). \end{aligned} \tag{6}$$

$$\begin{aligned} \tilde{C} &= \prod_{i=1}^{m} C_{GW_i} \cdot g^{\sum_{i=1}^{m} \tilde{m}} mod N^2 \\ &= (\prod_{i=1}^{m} (\prod_{j=1}^{n} g^{M_{ij}} \cdot r_{ij}^* mod N^2)) g^{\sum_{i=1}^{m} \tilde{m}} \\ &= (\prod_{j=1}^{n} (\prod_{i=1}^{m} g^{M_{ij}} \cdot r_{ij}^* mod N^2)) g^{\sum_{i=1}^{m} \tilde{m}} \\ &= (\prod_{j=1}^{n} (g^{M_{1j}} \cdot g^{M_{2j}} \cdot \ldots \cdot g^{M_{mj}}) (\prod_{i=1}^{m} \prod_{j=1}^{n} r_{ij}^* mod N^2)) g^{\sum_{i=1}^{m} \tilde{m}} \end{aligned} \tag{7}$$

Let $R^* = \prod_{i=1}^{m} \prod_{j=1}^{n} r_{ij}^*$. We have

$$\begin{aligned} \tilde{C} &= g^{\sum_{j=1}^{n} M_{1j} + \sum_{j=1}^{n} M_{2j} + \ldots + \sum_{j=1}^{n} M_{mj} + \sum_{i=1}^{m} \tilde{m}} R^* mod N^2 \\ &= g^{R_1^1 \sum_{j=1}^{n} \sum_{v=1}^{l} R^v d_{1jv} + \ldots + R_1^m \sum_{j=1}^{n} \sum_{v=1}^{l} R^v d_{mjv} + \sum_{i=1}^{m} \tilde{m}} R^* mod N^2 \\ &= g^{R_1^1 \sum_{j=1}^{n} R^v \sum_{v=1}^{l} d_{1jv} + \ldots + R_1^m \sum_{j=1}^{n} R^v \sum_{v=1}^{l} d_{mjv} + \sum_{i=1}^{m} \tilde{m}} R^* mod N^2 \end{aligned} \tag{8}$$

Let $B_{iv} = \sum_{j=1}^{n} d_{ijv}$,which represents the sum of the

electricity data of the first v dimension of all users in community i. We have $B_i = \sum_{v=1}^{l} R^v \cdot B_{iv}$ and $\tilde{m} = \sum_{i=1}^{m} \tilde{m}_i$.

Let

$$\begin{aligned} M &= R_1^1 \sum_{j=1}^{n} R^v \sum_{v=1}^{l} d_{1jv} + ... + R_1^m \sum_{j=1}^{n} R^v \sum_{v=1}^{l} d_{mjv} \\ &= \sum_{i=1}^{m} R_1^i \sum_{v=1}^{l} R^v \cdot B_{iv} \\ &= \sum_{i=1}^{m} R_1^i \cdot B_i \end{aligned} \tag{9}$$

We have $\tilde{M} = M + \tilde{m}$. Because $\tilde{C} = g^{\tilde{M}} \cdot r_{ij}^{*N} mod N^2$ is still *Paillier* encryption algorithm form, we can get $\tilde{M}$ by using the corresponding private key (λ, μ). Here, although $\tilde{M}$ is the noisy data, the impact of noise added temporarily can be ignored. It will not affect the operation center for its analysis, because the accuracy of the meter level allows the existence of the error and the error can be controlled within the allowable range by changing the size of ε.

By executing the data recovery algorithm [19] in Table 2, using $\tilde{M}$ and R_1 as the input of the algorithm, OC can obtain B_i. Using B_i and R as the input of the algorithm, OC can get B_{iv}.

Algorithm 1 Data recovery algorithm

Input:A and x $//A = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x$
Output:$\{a_1, a_2, ...a_n\}$

1: $X_0 = A / x$
2: **for** $j = 1$ to l **do**
3: $\quad a_j = X_{j-1} \bmod x$
4: $\quad X_j = X_{j-1} \bmod x$
5: **end for**
6: **return** $\{a_1, a_2, ...a_l\}$

6 Security Analysis

In this section, we will show that our multidimensional data aggregation scheme implements the security requirements that are proposed in Section 3.

- Confidentiality: The confidentiality of user data is protected. Because the users' encrypted data exist in the form of ciphertext $C_{ij} = g^{M_{ij}} \cdot r_{ij}^{*N} mod N^2$ during the transmission process, and the *Paillier* cryptography system is semantically secure for the selected plaintext attacks, the adversary can not obtain the user's electricity information. Even if the adversary invades the gateway database, he/she still can not get the user's specific electricity information. The malicious users who want to analyze other people's usage information may collude with each other to share their own data, but they can not infer electricity consumption of other users because the user's private key is a secret storage.

- Unforgeability and Data Integrity: The authentication and integrity of user data and aggregation data are protected. In this scheme, the user's private key is composed of two parts, one part is the user certificate generated by OC, and the other part is the secret value, which is independently selected by the user. Therefore, the security certification of certificate-based signature is to prove that only know the user's certificate and secret value (*ie*, fully aware of the user's private key) to produce a valid signature. If an adversary wants to crack a signer's private key or certificate, it faces the difficulty of solving the discrete logarithm, which is safe in the random prediction model under the CDH model [12, 23, 24].

- Differential Privacy: User data can resist differential attacks. Differential attack is to change the input of the algorithm, through the output of the algorithm or the change value of the output to expose the information in the algorithm input. If we use an algorithm to compute the aggregation result of a group of users' usage data, the differential attack can obtain a set of aggregation results by using the algorithm again. If the data of the two aggregations differ by only one user U, then the power consumption data of the user U can be obtained by subtracting the value of the two aggregation results.

  In this scheme, ε-differential privacy is achieved by adding noise to the community gateway with a given privacy level of ε. Even if the external adversary initiates a differential attack by analyzing the similar data sets obtained from the two aggregations, he/she gets the data with noise and can not calculate the electricity usage information of a single user.

  Assume that the adversary A obtains two perturbed aggregations $s + \tilde{m}_s$ and $t + \tilde{m}_t$, where s and t are two data sets with only one element different and $\tilde{m}_s$ and $\tilde{m}_t$ are two corresponding *Laplacian* noise. Because$|s - t| \leq W$, for any integer k, we have

$$\begin{aligned} \theta &= P(s + \tilde{m}_s = k)/P(t + \tilde{m}_t = k) \\ &= P(\tilde{m}_s = k - s)/P(\tilde{m}_t = k - t) \\ &= \frac{1}{2b} e^{-\frac{|k-s|}{b}} / \frac{1}{2b} e^{-\frac{|k-t|}{b}} \\ &= e^{-\frac{|k-s|-|k-t|}{b}} \end{aligned} \tag{10}$$

  Because$-|s-t| \leq |k-s|-|k-t| \leq |s-t|$, we have $\theta \leq e^{|s-t|/b} \leq e^{W/b} = e^{\varepsilon}$ satisfying ε-differential privacy. So the aggregated data set is added enough noise to provide differential privacy for each participating user while still provides high efficiency. Our scheme is safe in ε-differential privacy and can provide a strong and provable privacy guarantee.

Table 2: Comparison of five multidimensional electricity data aggregation schemes

Performance	Our scheme	Lu's scheme	Shen's scheme	Fu's scheme	Zhou's scheme
Confidentiality	Yes	Yes	Yes	Yes	Yes
Unforgeability	Yes	No	Yes	No	Yes
Signature verification security	Yes	No	Yes	No	Yes
Multi-level gateway	Yes	No	Yes	No	Yes
Dynamic users	Yes	No	Yes	No	Yes
Differential privacy	Yes	No	No	No	No

7 Performance Analysis

This section describes our scheme from the implemented functions, the computational cost, the communication overhead and the differential privacy utility. Table 2 shows the performance of the proposed scheme and the other four multidimensional data aggregation schemes [9, 17, 19, 26]. We compared confidentiality, data integrity, unforgeability, signature verification security, multi-level gateway, dynamic users and differential privacy. Lu's scheme [17] and Fu's scheme [9] have problems in security of batch authentication signatures. Although Fu's scheme [9] can handle the aggregation of multidimensional data, the aggregated result is not the fine-grained data of each dimension. Shen's scheme [19] solves the problem of counterfeiting in batch verification. But the above schemes can not resist the differential attacks.

7.1 Computational Complexity

In our scheme, C_e,C_m and C_p represent the computational cost of an exponentiation operation in Z_{n^2}, a scalar multiplication operation in G_1 and a pairing operation. According to Shen's scheme [19], the computational cost of an exponential operation, a pairing operation and a multiplication operation is shown in Table 3. Compared with the exponential operations and pairing operations, the computational cost of the multiplication operations in Z_n and the hash operations are considered negligible.

For our proposed scheme, a user U_{ij} needs to perform two exponential operations in Z_{n^2} to generate C_{ij} and a scalar multiplication operation in G_1 to generate the signature. A community gateway needs to perform three pairing operations and $2n_i$ scalar multiplication operations to verify the aggregated signature. An exponential operation needed to be used to add noise. A community gateway generates the signature with a scalar multiplication operation. In order to verify the aggregated data from the BGWs, a regional gateway need to perform three pairing operations and $2m$ multiplication operations. A DGW need to generate the signature with a scalar multiplication operation. For the OC, three pairing operations and two multiplication operations are used to verify the signature.

In Table 4 we compare the computational cost of our scheme with Shen's scheme [19], Zhou's scheme [26] and our scheme. The computational cost of the user in Zhou's scheme is related to the dimension of the data, but our scheme and Shen's scheme are independent of the dimension of the data. Obviously our scheme has a better performance than Zhou's scheme. Through the following analysis, we show that our scheme is better than Shen's scheme too.

On the community gateway side, we assume that the maximum number of users per community is $n(n \geq n_i)$, let

$$\begin{aligned} 3C_p + C_e + (2n+1)C_m &\leq (n+2)C_p + C_m \\ 78.8 + 12.8n &\leq 20n + 46.4 \\ 7.2n &\geq 32.4 \\ n &\geq 4.5 \end{aligned} \tag{11}$$

When the number of users in a community is greater than or equal to 5, our scheme has a less computational cost. From Figure 2, it can be seen our scheme has a better performance on the community gateway side.

On the regional gateway side, m is the number of community gateways, let

$$\begin{aligned} 3C_p + (2m+1)C_m &\leq (m+2)C_p + C_m \\ 66.4 + 12.8m &\leq 20m + 46.4 \\ 7.2m &\geq 20 \\ m &\geq 2.78 \end{aligned} \tag{12}$$

When the number of community gateways is greater than or equal to 3, our scheme has a less computational cost.

From Figure 3, it can be seen our scheme has a better performance on the regional gateway side.

According to Table 4, we compared the total computational cost of our scheme, Shen's scheme and Zhou's scheme. The computational cost of Zhou's scheme is related to the dimension of the data, so we assume the dimension of the data is 3. From Figure 4, Figure 5 and Figure 6, it is obvious that our scheme has a better performance.

7.2 Communication Overhead

For each user encrypted data $(ID_{ij}, \sigma_{ij}, C_{ij}, T)$, the communication overhead for all users in a community to the

Table 3: Description Calculation time (ms)

Symbol	Description	Calculation time (ms)
C_e	Exponential operation	12.4
C_p	Bilinear pairing operation	20
C_m	Multiplication	6.4

Table 4: The cost comparison

Symbol	Our scheme	Shen's scheme	Zhou's scheme
$User$	$2C_e + C_m$	$2C_e + C_m$	$(l+1)C_e + C_m$
BGW	$3C_p + (n - n_i + 1)C_e + (2n_i + 1)C_m$	$(n_i + 2)C_p + (n - n_i)C_e + C_m$	$2C_p + (4n - 1)C_m + C_e$
DGW	$3C_p + (2m + 1)C_m$	$(m + 2)C_p + C_m$	$2C_p + (4m - 2)C_m$
OC	$3C_p + 2C_m$	$2C_p$	$2C_p$

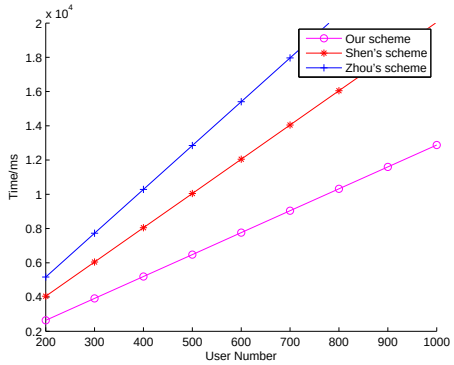

Figure 2: The computational cost of each BGW

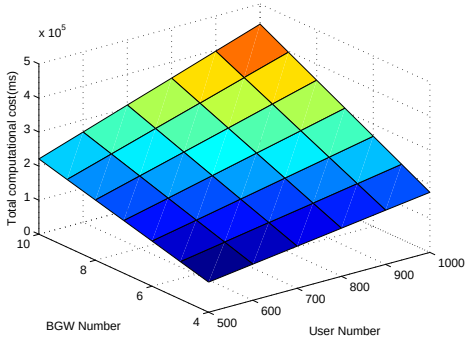

Figure 4: The total computational cost of our scheme

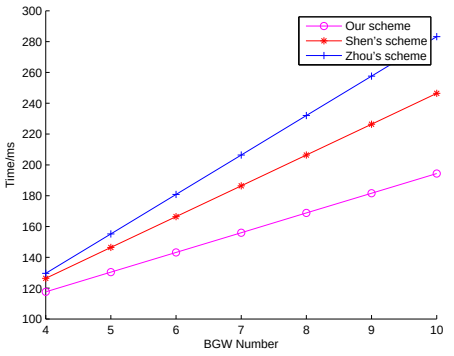

Figure 3: The computational cost of each DGW

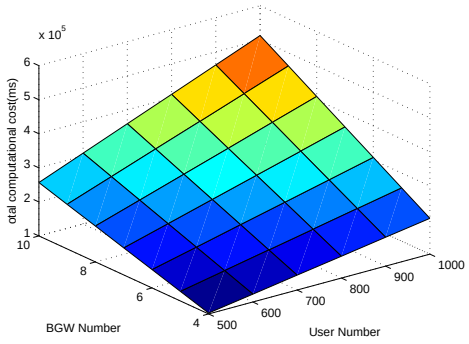

Figure 5: The total computational cost of Shen's scheme

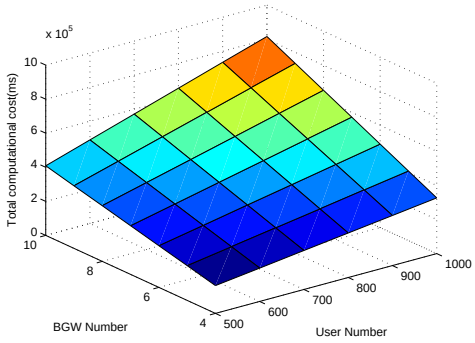

Figure 6: The total computational cost of Zhou's scheme

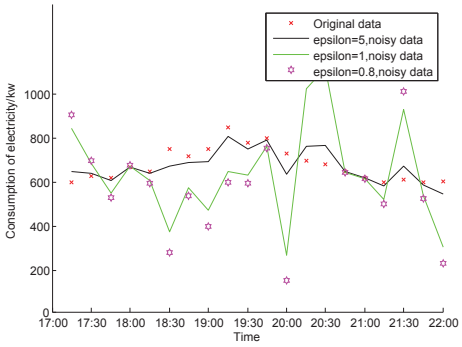

Figure 7: Differential privacy utility comparison

community gateway is $S_{U\to BGW} = (|ID_{ij}|+|\sigma_{ij}|+|C_{ij}|+|T|)n$. For each community gateway aggregated data $(ID_i, \sigma_i, \tilde{C}_i, T)$, the communication overhead for all community gateways in a region to the community gateway is $S_{BGW\to DGW} = (|ID_i| + |\sigma_i| + |\tilde{C}_i| + |T|)m$. For each regional gateway aggregated data $(ID, \sigma, \tilde{C}, T)$, the communication overhead for a regional gateway to the operations center is $S_{DGW\to OC} = |ID| + |\sigma| + |\tilde{C}| + |T|$.

We choose 1,024 bits N ($|N^2| = 2048$) and 160 bits G_1, set $|ID|+|T|$ to 64 bits. Table 5 shows the communication overhead comparison of our scheme and other two schemes [17, 19].

The total communication cost of our scheme is $2272 * m * n + 2272 * m + 2272$ and the total communication cost of Shen's scheme is $2308 * m * n + 2308 * m + 2308$. Our scheme has less communication overhead than the other schemes.

7.3 Differential Privacy Utility Comparison

In order to verify the effectiveness of differential privacy, we assume that a community has 1,000 users. Between 17: 00-22: 00, every 15 minutes of electricity data is aggregated, the community gateway adds Laplace noise to achieve ε - differential privacy. Through the inverse cumulative distribution function of the Laplace distribution, we obtain the Laplace noise by inputting the random variables uniformly distributed in the range of (-0.5, 0.5) for this inverse cumulative distribution function.

ε is the privacy budget, based on the data,we have $\Delta f = 200$.

From Figure 7, it can be seen that the smaller ε we use , the better the effect of privacy protection we have, but the utility is relatively poor. The large ε we use, the utility is better. The literature [14] made a more detailed introduction about how to choose ε.

8 Conclusions

In this paper, a secure multidimensional data aggregation scheme is proposed, which uses the Horner rule to deal with polynomials of multidimensional electricity data. The method of certificate aggregation is used to realize the unforgeability of authentication and signature. Time-consuming pairing is reduced to a constant; by adding Laplace noise to the community gateway to achieve ε - differential privacy to resist differential attacks; through performance analysis, the scheme's computational and communication overhead is improved.

Acknowledgments

This work was supported by NSFC Grants (No.61772327 No.61202020 No.61532021), Project of Shanghai Science and Technology Committee Grant (No.15110500700) and CCF-Tencent Open Fund Grant (No.IAGR20150109, RAGR20150114). We would like to express our gratitude to the anonymous reviewers for their valuable feedback and comments which helped us to improve the quality and presentation of this paper.

References

[1] A. Abdallah and X. Shen, "Lightweight security and privacy preserving scheme for smart grid customer-side networks," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1064–1074, 2017.

[2] M. Badra and S. Zeadally, "An improved privacy solution for the smart grid," *International Journal of Network Security*, vol. 18, pp. 529-537, 2016.

[3] H. Bao and R. Lu, "A lightweight data aggregation scheme achieving privacy preservation and data integrity with differential privacy and fault tolerance," *Peer-to-Peer Networking and Applications*, vol. 10, no. 1, pp. 1–16, 2015.

[4] P. Barbosa, A. Brito and H. Almeida, "A technique to provide differential privacy for appliance usage in

Table 5: Comparison of communication overhead

	Our scheme	**Shen's scheme**	**Lu's scheme**
$U \rightarrow BGW$	2272	2308	2308
$BGW \rightarrow DGW$	2272	2308	
$DGW \rightarrow OC$	2272	2308	2308

smart metering," *Information Sciences*, vol. 370-371, pp. 355–367, 2016.

[5] L. Chen, R. Lu, Z. Cao, K. Alharbi and X. Lin, "Muda: Multifunctional data aggregation in privacy-preserving smart grid communications," *Peer-to-Peer Networking and Applications*, vol. 8, no. 5, pp. 1–16, 2014.

[6] C. Dwork, "Differential privacy: A survey of results," in *International Conference on Theory and Applications of MODELS of Computation*, pp. 1–19, 2008.

[7] C. Dwork, F. Mcsherry and K. Nissim, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography Conference*, pp. 265–284, 2006.

[8] C. I. Fan, S. Y. Huang and Y. L. Lai, "Privacy-enhanced data aggregation scheme against internal attackers in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 666–675, 2013.

[9] S. Fu, J. Ma, H. Li and Q. Jiang, "A robust and privacy-preserving aggregation scheme for secure smart grid communications in digital communities," *Security & Communication Networks*, vol. 9, no. 15, pp. 2779–2788, 2016.

[10] D. He, N. Kumar and J. H. Lee, "Privacy-preserving data aggregation scheme against internal attackers in smart grids," *Wireless Networks*, vol. 22, no. 2, pp. 491-502, 2016.

[11] D. He, S. Zeadally, H. Wang and Q. Liu, "Lightweight data aggregation scheme against internal attackers in smart grid using elliptic curve cryptography," *Wireless Communications & Mobile Computing*, vol. 2017, pp. 1–11, 2017.

[12] L. Huimin, L. Hongmei, W. Haimin and Z. Jinhui, "An efficient certificate-based aggregate short scheme," *Journal of Ningxia University(Nature Science Edition)*, no. 1, pp. 52–57, 2017.

[13] K. Kursawe, G. Danezis and M. Kohlweiss, "Privacy-Friendly Aggregation for the Smart-Grid", in *International Symposium on Privacy Enhancing Technologies Symposium*, pp. 175-191, 2011.

[14] J. Lee and C. Clifton, "How much is enough? choosing e for differential privacy," in *International Conference on Information Security*, pp. 325–340, 2011.

[15] F. Li, B. Luo and P. Liu, "Secure and privacy-preserving information aggregation for smart grids," *International Journal of Security & Networks*, vol. 6, no. 1, pp. 28–39, 2011.

[16] H. Li, X. Lin, H. Yang, X. Liang, R. Lu and X. Shen, "Eppdr: An efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid," *IEEE Transactions on Parallel & Distributed Systems*, vol. 25, no. 8, pp. 2053–2064, 2014.

[17] R. Lu, X. Liang, X. Li, X. Lin and X. Shen, "Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Transactions on Parallel & Distributed Systems*, vol. 23, no. 9, pp. 1621–1631, 2012.

[18] S. Nath and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *ACM SIGMOD International Conference on Management of Data*, pp. 735–746, 2010.

[19] H. Shen, M. Zhang and J. Shen, "Efficient privacy-preserving cube-data aggregation scheme for smart grids," *IEEE Transactions on Information Forensics & Security*, vol. 12, no. 6, pp. 1369–1381, 2017.

[20] R. Singh and M. S. Manu, "An energy efficient grid based static node deployment strategy for wireless sensor networks," *International Journal of Electronics and Information Engineering*, vol. 7, no. 1, pp. 32-40, 2017.

[21] X. Tian, L. Li, J. Li, H. Li, and C. Gu, "Secret share based program access authorization protocol for smart metering," *International Journal of Network Security*, vol. 18, no. 6, pp. 1071–1079, 2016.

[22] S. F. Tzeng, M. S. Hwang, "Digital signature with message recovery and its variants based on elliptic curve discrete logarithm problem", *Computer Standards & Interfaces*, vol. 26, no. 2, pp. 61–71, Mar. 2004.

[23] J. Xu, Z. Zhang and D. Feng, "Id-based aggregate signatures from bilinear pairings," *Lecture Notes in Computer Science*, vol. 3810, pp. 110–119, 2005.

[24] H. Xiong, Z. Qin and F. Li, "Identity-based threshold signature secure in the standard model," *International Journal of Network Security*, vol. 10, no. 1, pp. 75–80, 2010.

[25] T. xiuxia, L. lisha, S. Chaochao and L. D. Ming, "Review on privacy protection approaches in smart meter," *Journal of East China Normal University(Nature Science)*, no. 5, pp. 46–60, 2015.

[26] H. Zhou, J. Chen, Y. Y. Zhang and L. J. Dang, "A multidimensional data aggregation scheme in multi-level network in smart grid," *Journal of Cryptologic Research*, vol. 4, no. 2, pp. 114–132, 2017.

Biography

Xiuxia Tian Professor, received the MS degree in applied cryptography-based information security from Shanghai Jiaotong University in 2005, and the PhD degree in database security and privacy preserving in cloud computing from Fudan University in 2011. She is currently a professor in the College of Computer Science and Technology, Shanghai University of Electric Power. She is a visiting scholar of two years at UC Berkeley working with groups of SCRUB and SecML. She has published more than 40 papers and some papers are published in international conferences and journals such as DASFAA, ICWS, CLOUD, and SCN. Her main research interests include database security, privacy preserving (large dataand cloud computing), applied cryptography, and secure machine learning.

Qian Song Graduate, College of Computer Science and Technology in Shanghai University of Electric Power. Her research interests mainly focus on the security and privacy protection for the smart meter.

Fuliang Tian Graduate, Graduate. College of Computer Science and Technology in Shanghai University of Electric Power. He research interests mainly focus on the security and privacy protection for the smart meter. Email: tianflxs@163.com.